HDP Developer: Apache Pig and Hive (Hadoop)

Learn via: **Classroom / Virtual Classroom / Online**

Duration: **4 Gün**

https://bilginc.com/tr/egitim/hdp-developer-apache-pig-and-hive-hadoop-941-egitimi/

Overview

Bu eğitim, Pig ve Hive kullanarak Apache Hadoop'da saklanan Büyük Verileri analiz edecek uygulamaları geliştirmesi gereken geliştiriciler için tasarlanmıştır. Hadoop, YARN, HDFS, MapReduce, veri özümseme, iş akışı tanımı, Big Data veri analitiği için Pig ve Hive'ı kullanma ve Spark Core ile Spark SQL'e giriş konularını içerir.

Who Should Attend

Hadoop için uygulamaları anlaması ve geliştirmesi gereken yazılım geliştiricileri bu eğitime katılabilir.

What You Will Learn

Bu eğitimin tamamlanmasıyla birlikte öğrenciler:

- Hadoop'u, YARN'ı ve Hadoop kullanım vakalarını açıklayabilecekler
- Hadoop ekosistem araçlarını ve çalışma çerçevelerini açıklayabilecekler
- HDFS mimarisini açıklayabilecekler
- HDFS'ye veri girmek için Hadoop istemcisini kullanabilecekler
- Hadoop ile bir ilişkisel veritabanı arasında veri transferi yapabilecekler
- YARN ve MapReduce mimarilerini anlatabilecekler
- YARN üzerinde bir MapReduce işini çalıştırabilecekler
- HDFS'de verileri keşfetmek ve dönüştürmek için Pig kullanabilecekler
- Hive tablolarının nasıl tanımlandığını ve uygulandığını anlayabilecekler
- Veri kümelerini keşfetmek ve analiz etmek için Hive'ı kullanabilecekler
- ORC dosya formatlarını kullanan bir Hive tablosunu oluşturabilecek ve doldurabilecekler
- Çeşitli Hive dosya formatlarını anlatabilecek ve kullanabilecekler
- Yeni Hive pencereleme fonksiyonlarını kullanabilecekler
- Veri analizi için SQL benzeri sorguları çalıştırmak üzere Hive'ı kullanabilecekler
- Çeşitli teknikleri kullanarak veri kümelerini birleştirmek için Hive'ı kullanabilecekler
- Etkin Hive sorguları yazabilecekler
- DataFu Pig kitaplığını kullanarak veri analitiği gerçekleştirebilecekler
- HCatalog kullanımlarını ve amacını anlatabilecekler
- Pig ve Hive ile HCatalog'u kullanabilecekler
- Bir Oozie iş akışını tanımlayabilecek ve planlayabilecekler
- Spark ekosistemini ve üst düzey mimarisini sunabilecekler
- Spark'ın Resilient Distributed Dataset API'si ile veri analizi yapabilecekler
- Spark SQL ve DataFrame API'yi keşfedebilecekler

Outline

- Use HDFS commands to add/remove files and folders
- Use Sqoop to transfer data between HDFS and a RDBMS
- Run MapReduce and YARN application jobs
- Explore, transform, split and join datasets using Pig
- Use Pig to transform and export a dataset for use with Hive
- Use HCatLoader and HcatStorer
- Use Hive to discover useful information in a dataset
- Describe how Hive queries get executed as MapReduce jobs

- Perform a join of two datasets with Hive
- Use advanced Hive features: windowing, views, ORC files
- Use Hive analytics functions
- Write a custom reducer in Python
- Analyze clickstream data and compute quantiles with DataFu
- Use Hive to compute ngrams on Avro-formatted files
- Define an Oozie workflow
- Use Spark Core to read files and perform data analysis
- Create and join DataFrames with Spark SQL